



Review

Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations [☆]

István Winkler ^{a,b,*}, István Czigler ^{a,c}^a Institute for Psychology, Hungarian Academy of Sciences, Budapest, Hungary^b Institute of Psychology, University of Szeged, Szeged, Hungary^c Institute for Psychology, Eötvös Loránd University, Budapest, Hungary

ARTICLE INFO

Article history:

Received 17 July 2011

Received in revised form 3 October 2011

Accepted 5 October 2011

Available online 30 October 2011

Keywords:

Perception

Predictive coding

Deviance detection

Object representation

Event-related brain potentials (ERP)

Mismatch negativity (MMN)

Visual mismatch negativity (vMMN)

ABSTRACT

Predictive coding theories posit that the perceptual system is structured as a hierarchically organized set of generative models with increasingly general models at higher levels. The difference between model predictions and the actual input (prediction error) drives model selection and adaptation processes minimizing the prediction error. Event-related brain potentials elicited by sensory deviance are thought to reflect the processing of prediction error at an intermediate level in the hierarchy. We review evidence from auditory and visual studies of deviance detection suggesting that the memory representations inferred from these studies meet the criteria set for perceptual object representations. Based on this evidence we then argue that these perceptual object representations are closely related to the generative models assumed by predictive coding theories.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Helmholtz's (1860/1962) notion of unconscious inference engendered arguably the most fruitful line of perceptual research throughout the relatively short history of psychology, the empiricist tradition. In one of its contemporary variants, Gregory (1980) suggests that perception is akin to scientific hypotheses: it is the brain's best-fitting model for the information entering the senses. But together with Gordon (1997) we can ask how these models are formed, what evidence they are tested against, and how they adapt to an ever changing environment? To answer these questions, some of the theories of predictive coding (Creutzig and Sprekeler, 2008; Dayan et al., 1995; Friston, 2005, 2010; Hohwy, 2007; Mumford, 1992; Rao and Ballard, 1999; Schütz-Bosbach and Prinz, 2007) evoke the principle of free-energy minimization (e.g., Friston, 2005, 2010).

Predictive coding theories suggest that the perceptual system's primary objective is to minimize the discrepancy between predictions from its internal generative models of the environment and the actual

sensory input. Structured as a hierarchy of models of increasing levels of abstraction, predictions from each level are tested on data emerging one level lower with the difference (termed the "error signal" or "prediction error") being passed upwards in the hierarchy (for non-mathematical descriptions, see Baldeweg, 2007; Hohwy et al., 2008). The error signal then governs model selection/adjustment in order to minimize prediction error throughout the system. Thus predictive coding theories implement the analysis by synthesis principle (Neisser, 1967; Yuille and Kersten, 2006) and conform to the notion of gist-first processing suggested by some recent theories of perception (Ahissar and Hochstein, 2004; Bar, 2004, 2007), whereby higher-level (more general) models govern the interpretation (model selection) at lower levels. Predictive coding theories acknowledge the stochastic nature of the information entering the senses, a notion that has long been argued by an early theorist of perception, Egon Brunswik (1956). Dealing with probability distributions instead of discrete values, predictive coding theories assume that the brain follows Bayesian inference rules in model selection (Kersten et al., 2004; Knill and Pouget, 2004; Yuille and Kersten, 2006). Models based on hierarchical Bayesian inference using hierarchical generative models represent a recent development in the field (Friston and Kiebel, 2009; Lee and Mumford, 2003). In a hierarchical setting, the predictions from higher levels play the role of empirical priors on representations in lower levels. This resolves concerns about where priors come from and makes (empirical) priors accountable to

[☆] Contribution to the Special Issue titled "Predictive information processing in the brain: Principles, neural mechanisms and models" edited by J. Todd, E. Schröger, and I. Winkler.

* Corresponding author at: Institute for Psychology, Hungarian Academy of Sciences, H-1394 Budapest, P.O. Box 398., Hungary. Tel.: +36 1 3542296; fax: +36 1 3542416. E-mail address: iwinkler@cogpsyphy.hu (I. Winkler).

sensory data. Thus sensory data is used to update the evaluation (the probability of the correctness) of existing models. In the end, the model with the highest probability of being correct determines the (conscious) percept.

Thus, according to these theories, the general makeup of the afferent system¹ is divided into 1) neuronal circuits implementing the generative models and setting up lower levels in the hierarchy and 2) circuits determining prediction errors and passing them onto higher levels (Friston, 2005). However, whereas the determination of prediction errors is quite clear, the make-up of the corresponding generative models is rather unspecified beyond the principles of Bayesian inference processing. A consequence of this imbalance of detail between the two assumed functional units of predictive coding theories is that most neuroscience evidence interpreted in favor of predictive coding comes from observing neuronal activity that shows effects expected of processing prediction errors. The major sources of such evidence are single-cell data and simulations (Grill-Spector et al., 2006; Hosoya et al., 2005; Jehee and Ballard, 2009; Lee and Mumford, 2003; Wang et al., 2006), local field potentials (Kumar et al., in press), and large-scale brain responses (Alink et al., 2010; Aoyama et al., 2005; den Ouden et al., 2010; Murray et al., 2002); each showing reduced activity for predicted as compared with unpredicted sensory input. There exist also behavioral data compatible with what is expected from a system working on Bayesian principles (den Ouden et al., 2010; Ernst and Banks, 2002; Hohwy et al., 2008; Weiss et al., 2002; Yu, 2007). However, the representation and maintenance of the generative models received less elaboration so far.

Psychological theories agree on that the overall function of perception is to discover the sources of the information entering the senses, because knowledge about these objects and events can be utilized to reach survival and reproduction goals (e.g., Brunswik, 1956). Thus behavior is influenced by the distal objects and events. Even when behavior is apparently controlled by a single feature (e.g. we pick up a cherry by its color), the feature belongs to an object. Therefore, psychological theories have for a long time assumed the existence of brain representations for objects and suggested that incoming sensory information is stored and manipulated in such units in the brain. The question addresses here is how these representations relate to the multi-leveled generative models of predictive coding theories?

The representations inferred from studies measuring the mismatch negativity (MMN: Näätänen et al., 1978, for a recent review, see Näätänen et al., 2011) event-related brain potential (ERP) and its visual counterpart (vMMN: Tales et al., 1999 and Heslenfeld, 2003, for recent reviews, see Czigler, 2007, 2010, and Kimura, in press in this issue) may provide a useful link between these two views of perception. MMN and vMMN are elicited when the incoming stimulus violates some regular feature detected from the preceding sequence. MMN was discovered in the context of the auditory oddball paradigm. Occasionally exchanging a repetitive sound (termed, the “standard”) for a different one (termed, the “deviant”) elicited a fronto-centrally negative ERP response (MMN) peaking between 100 and 200 ms from the onset of the deviance (typically the sound onset). MMN was initially described as an ERP correlate of detecting a mismatch between the memory trace of the repeating sound and that of the incoming one (Näätänen et al., 1978). Research in the past thirty years demonstrated that MMN is also elicited by violations of regularities which are more complex than stimulus repetition, including such regularities in which each sound is specified by the immediately preceding one (Paavilainen et al., 2007; Horváth et al., 2001). These and similar evidence as well as a detailed analysis of

the alternative interpretations (see Winkler, 2007) led to the hypothesis that 1) memory representations of the detected regularities are generative models providing predictions about upcoming sensory events and 2) MMN is elicited when the current stimulus does not match these predictions (Baldeweg, 2006, 2007; Garrido et al., 2009c; Näätänen et al., 2011; Sinkkonen, 1999; Winkler, 2007; Winkler et al., 1996; see also Bendixen et al., in press in this issue). Winkler and Czigler (1998) further argued that the function of the MMN signal is to update the regularity representations violated by the deviant stimulus (see also Winkler, 2007). Thus, in terms of predictive coding theories, MMN can be regarded as a signal carrying the prediction error (Garrido et al., 2009c).

Based on the above interpretation of MMN, the memory representations reflected in the MMN ERP component may be compatible with the generative models assumed in predictive coding descriptions of perception. We previously suggested (Winkler, 2010; Winkler et al., 2009) that the representations inferred from MMN studies meet the criteria set for auditory object representations. Thus results obtained in studies of the auditory and visual MMN may provide a link between the predictive coding view of perception and the psychological literature of perceptual object representations.

Here we review results of studies measuring the auditory and visual MMN offering (indirect) evidence about the nature of perceptual object representations. The aims of the review are 1) to compare characteristics of the object representations in the two modalities and 2) to assess how well they fit into a generalized predictive coding account of perception.²

2. Object representations and (v)MMN

Objects serve as perceptual units, as was first emphasized by Gestalt psychologists (Köhler, 1947) and they are also the units of attentional selection (e.g., Duncan, 1984). The first difference between the two (auditory and visual) modalities lies in what constitutes this unit of representation. That is, what is a perceptual object? Whereas in vision, object representations unequivocally refer to physical objects in the environment, in the auditory modality, two different perceptual units can be distinguished. The acoustic source (the physical cause of the sounds) is termed “concrete object”, whereas the pattern of sounds (e.g., a melody or the sound of and approaching car) is termed “abstract object” (Wightman and Jenison, 1995). Although it may initially seem attractive to restrict the notion of auditory perceptual objects to concrete (physical) objects, it appears that abstract objects are just as, perhaps even more relevant in describing auditory perception (for detailed discussion, see Griffiths and Warren, 2004 and Kubovy and Van Valkenburg, 2001). The difference between the representational units in audition and vision stems from the physical nature of these modalities (for a discussion, see Bregman, 1990) and, consequently, the way humans utilize information emerging in them.³ Adapted to our natural environment and mode of acquiring food, the human perceptual system relies more on vision than audition for characterizing physical objects (their location, shape, make-up, etc.; see e.g., Spence and Driver, 1999). In contrast, fast processing of auditory information on several time scales in parallel provides early warning of danger and the means of communication (see e.g., Boemio, et al., 2005; Poeppel, 2003). These purposes are better supported by representations of abstract than concrete objects which allow one to select and recognize sound patterns characterizing various environmental events and communicative acts. Accepting abstract auditory objects as perceptual units forced theoreticians to describe object representations in

¹ By most predictive coding theories, motor control and cognition works similarly. We shall, however, restrict our discussion to perception.

² For links between MMN and Friston's influential predictive coding model of perception, see Garrido et al. (2009c).

³ Note that some species, such as bats, use auditory information similarly to how visual information is used by humans (through sound emission and echolocation).

terms of processing principles applicable across modalities, rather than attempting to find direct analogies between auditory and visual features (Gregory, 1980; Griffiths and Warren, 2004; Kubovy and Van Valkenburg, 2001; Winkler et al., 2009). In the following, when discussing object representations, we mean both types of auditory objects.

Subjectively, objects appear to us as the combination of their features (such as color, shape, size, etc. in the visual and pitch, timbre, duration, etc. in the auditory modality). Feature combinations are assumed to be represented together in what has been termed object files (e.g., Kahneman et al., 1992). However, this phenomenological description of object representations is not sufficient. For example, perceptual objects preserve their identity even when some of their features change. Feature changes may not be intrinsic to the object. They could result from changes in the environment surrounding the object (such as a change of the intensity of the illumination) or in the observer (e.g., a change in his/her position relative to the object). Mechanisms of perceptual constancies (for a detailed discussion, see e.g., Gordon, 1997) separate true changes of object features from changes due to environmental events, thus preserving object identity even when sensory information changes. In fact, the empiricist tradition as well as the roots of predictive coding lies in Helmholtz's (1860/1962) explanation of perceptual constancy in the face of environmental effects: He assumed that the brain utilizes predictive relationships between characteristics of the environment and the sensory information arriving from objects within it.

There are, however, also 'true' changes in object features. These are especially important when dealing with living entities. For example, change in color can signal when some food becomes edible, changes in the posture and voice of an animal can suggest that it prepares for an attack, etc. Typically, a set of features changes together following some regular pattern (e.g., the appearance of various emotions on the face and in the voice of a person). Being able to predict what follows these changes while preserving the identity of the object is of obvious survival value. Furthermore, object representations probably also involve non-perceptual features, such as the possible usage of the object, termed affordance (Gibson, 1977) or object-related actions (Hommel et al., 2001; Prinz, 1990), as well as attitudes toward the objects. Predictive coding theories with their hierarchically structured representations of increasing generality and complexity are especially well suited for providing an explanatory framework for maintaining object identity despite (predictable) transformations and to link perception with actions and cognitive constructs.

In this review, we discuss perceptual object representations based on information from a single modality. There are two main reasons to focus on such representations. Firstly, since the time of Gestalt psychology, a large part of the perceptual literature deals with this type of representations, gathering considerable knowledge about them. Thus we will be able to characterize them and examine how they fit into the larger picture provided by predictive coding theories. Secondly, unimodal perceptual representations form the level at which information originating from different modalities may be of compatible structure and information content, such that they can be systematically aligned and combined into multi-modal object descriptions.⁴ One important aim of the current review is to examine how similar (compatible) such representations are in the auditory and the visual modality as assessed via a common testing method, the deviance detection paradigm.

⁴ This is not to deny that there may be direct cross-modal effects at low levels of the sensory hierarchy. There is ample anatomical and physiological evidence suggesting the existence of such cross-modal interactions. However, even if one does not assume a strict modular view of perception, it appears likely that at low levels of the perceptual system, information processing is largely modality specific.

2.1. Features expected of perceptual object representations

Here we generalize the five defining characteristics proposed for auditory object representations by Winkler et al. (2009) to also cover visual object representations and link it with the principles of predictive coding theories.

- 1) Perceptual object representations encode and thus predict the combination of sensory features as they appear in the input. This feature is identical to the phenomenological description of perceptual object representations (see, e.g., Kahneman et al., 1992).
- 2) Perceptual object representations allow separation of different objects by specifying (predicting) which parts of the sensory input belong to them.
These two features of perceptual object representations do not take into account that the sensory input changes in time. However, it is highly unlikely that these features would be implemented in the human brain separately from those that deal with the effects of time. Thus these features are likely to emerge within a representation structure adapted to handle the sensory input as it unfolds in time.
- 3) Perceptual object representations connect the information belonging to the same object as it appears in time, thus predicting how information from an object becomes available in the future. Although most visual studies of object representations presented still pictures, in everyday visual scenes, parts of objects typically become visible at different times. In the auditory modality, abstract objects always represent temporal patterns, including intermittently appearing ones, such as a series of footsteps. Much information can only be derived by relating to each other samples from different times (e.g., whether the footsteps are approaching or receding; for a visual example consider an animal running through a wooded area).
- 4) Perceptual object representations absorb variation in the input (included in the generated predictions), while maintaining the identity of the represented object. The notion of identity goes beyond perception in the strict sense and likely involves multiple levels of representations. Variation resulting from ecologically valid changes in the environment or the observer may be handled at relatively low levels of the perceptual system. They are likely to be encoded in the initial structure of the sensory systems (representing learning through evolution⁵) and/or be part of the calibration of these systems by early sensory experience (such as those acquired by infants through manipulating objects). The results of other transformations, in terms of object identity, are highly sensitive to high-level knowledge about the context. Here we shall restrict the application of this criterion to the "primitive" perceptual constancies.
- 5) Perceptual object representations extend the actually available information about the object to parts that are currently not sensed (i.e., predictions are context sensitive). In the visual modality, typically not all parts of an object can be seen at any given moment of time. As Gregory (1980) argues, we perceive tables with four legs, even though we seldom see more than three of them at any given time. If perceptual object representations describe objects as they appear in time, then as we suggested above, they also extend the object information to future states. In fact, a functional view of perception needs to acknowledge that perception must be intrinsically directed toward the future, because we can only interact with future states of the environment compared with the time from which our information originated. For example, when crossing a street, we are not primarily concerned about where cars are at the moment, but where they will be when we reach their lane.

⁵ Note that the offspring of several species are born with highly functional perceptual (e.g., visual) systems and, at least initially, do not need or have the opportunity for extensive development through learning.

Some theorists suggest that the brain considers sensory information as encoding descriptions of future states of and events in the world (Creutzig et al., 2009; Dubnov, 2008; Tishby and Polani, 2011). Thus we consider extrapolation in general (cf. Gregory, 1980) and prediction of the future in particular as a defining feature of perceptual object representations (cf. Bar, 2007; Bubic et al., 2010; Kotz and Schwartz, 2010; Schütz-Bosbach and Prinz, 2007; Summerfield and Egner, 2009; Winkler, 2010; Winkler et al., 2009).

In the following subsections, we review the relevant MMN and vMMN studies for each of these characteristics of perceptual object representations.

2.2. Representing feature combinations

Results of MMN studies suggest that the memory representations involved in deviance detection encode the feature combinations of frequently encountered sounds (Winkler et al., 1990). Directly testing this issue, Gomes et al. (1997) composed sequences from three frequent ($p = .30$, each) and one infrequent ($p = .1$) pure tone. Each of the three frequent tones had a unique pitch and intensity level. The infrequent tone had the pitch of one and the intensity of another of the frequent tones (“conjunction deviant”). Thus the features of the conjunction deviant appeared frequently within the sequence, but the combination of these features was rare. MMN was elicited by these conjunction-deviant tones. This result, which has been since replicated by several studies (Sussman et al., 1998a; Takegata et al., 1999, 2001, 2005; Winkler et al., 2005a), shows that feature-conjunctions are encoded in the memory representations involved in auditory deviance detection.

A visual variant of Gomes et al.'s (1997) study was implemented by Winkler et al. (2005a). Square-wave grating stimuli were defined by two features, color (red/black or blue/black) and orientation (horizontal or vertical). Two of the color-orientation combinations appeared frequently ($p = .45$ each), whereas the other two infrequently ($p = .05$ each) within the sequences. Comparing between responses elicited by identical stimuli when they appeared infrequently vs. when they appeared frequently within the stimulus sequence, a vMMN was observed. Thus the memory representations involved in visual deviance detection also encode the feature combinations of the frequently presented stimuli. It should be noted that participants were unaware, which conjunctions were frequent and which conjunctions were infrequent. The automatic nature of the vMMN is supported by recent findings (Berti, 2011).

The literature on perceptual object formation is divided on the issue whether focal attention is required for representing feature combinations, that is, to create objects files. Basing on a wide range of behavioral evidence, Treisman and her colleagues (Treisman, 1998; Treisman and Gelade, 1980) suggested that features are only conjoined for those parts of the visual display to which spatial attention has been directed. In contrast, other researchers provided evidence suggesting that feature integration does not require focused attention (Duncan and Humphreys, 1989; Holcombe and Cavanagh, 2001; Houck and Hoffman, 1986). MMN (Takegata et al., 2005; Winkler et al., 2005a) and vMMN (Winkler et al., 2005a) results suggested that the features are correctly conjoined in both modalities outside the focus of attention. These results apparently contrast findings of illusory feature conjunctions found for stimuli falling outside the focus of attention (e.g., when observers do not have enough time to separately focus on each presented object) (in vision, Treisman and Schmidt, 1992; in audition Takegata et al., 2005; Thompson et al., 2001). Takegata et al. (2005) suggested that task-related strategies may be responsible for the reported illusory feature combinations. In contrast, the processing of task-irrelevant stimuli, as inferred from the MMN and vMMN results, is not biased by such top-down effects.

MMN studies also showed that a common feature is extracted from a sequence of sounds varying in other features (Deacon et al., 1998; Huotilainen et al., 1993; Gomes et al., 1995; Näätänen et al., 2004; Pakarinen et al., 2010; Winkler et al., 1990). To account for these findings as well as the above reviewed ones showing that feature-combinations are registered by the memory representations involved in MMN generation, Ritter et al. (1995) suggested that features and feature-combinations are stored in parallel. However, individual features are probably not processed independently of each other as some studies showed that rare sounds deviating from a repeating sound in multiple features may only elicit a single MMN (Czigler and Winkler, 1996; Oceák et al., 2008; Winkler and Czigler, 1998). For example, rare tones differing from the frequent ones in both pitch and duration, elicited an MMN response that was identical to the one elicited by pitch-violation alone, even though duration deviance also elicited the MMN response when the duration-deviant tone had the standard pitch (Czigler and Winkler, 1996). Sulykos and Czigler, (2011) provided similar evidence for the visual modality. Task-irrelevant Gábor-patches were presented regularly to participants. In separate stimulus sequences, occasionally, stimuli deviating in orientation, or spatial frequency, or both features (“double deviants”) were delivered. Double-deviants elicited a response that was identical to the orientation-only vMMN component. The absence of vMMN to spatial-frequency deviance within the double-deviants indicated that visual deviance detection was also based on integrated stimulus events, rather than separate features. Thus it appears that in both modalities, the feature-conjoined representation takes precedence over individual features in deviance detection.

Finally two or more features appearing together are not the only inter-feature relationship that is detected by the human auditory system. Paavilainen et al. (2001) found evidence that also rules, such as “the higher the pitch, the lower the intensity” were extracted, because tones violating this rule elicited the MMN.

Thus we conclude that the memory representations inferred from MMN and vMMN results encode the relationship between (in most cases, the combination of) the features of the stimuli appearing in a sequence.

2.3. Object separation

Segregation of concurrent sound objects can be based on instantaneous and/or temporal/sequential cues (Bregman, 1990). Instantaneous segregation utilizes the harmonic (spectral) structure of the sound input, relative timing of sound onsets, etc. Importantly, it does not require information regarding the preceding auditory stimulation and thus needs no predictive processing. The object related negativity (ORN) is regarded as the footprint of instantaneous sound segregation in ERPs (for a review, see Alain and Winkler, in press).

Segregation by temporal/sequential cues can be studied using MMN by making the regularities emerging in a stimulus sequence dependent on the perceptual organization of the sequence. This method is based on the observation that linking two sounds within a single object is much easier than linking two sounds belonging to separate objects. For example, when sounds are perceptually sorted into two coherent sequences (termed auditory streams) observers have difficulties to determine the relative timing or even the order between sounds belonging to different streams (Bregman, 1978; Bregman and Campbell, 1971). Several studies showed that violating sequential regularities led to MMN elicitation only when the parameters of the sound sequence promoted an organization in which the violated regularity could be perceived (e.g., Sussman et al., 1998b; Winkler et al., 2003; Yabe et al., 2001; for reviews, see Alain and Winkler, in press; Sussman, 2007; Winkler, 2010). For example, Sussman et al. (1998b) delivered a sequence of alternating high and low tones. Separately within the high and low tones, a three-tone pattern was cyclically repeated. Occasional reversals of the order of the three tones

(separately for the high and low pattern) elicited the MMN when the sequence was perceptually segregated into a high and a low stream, but not when all tones were heard as part of a single stream. This was because the single integrated stream was perceptually dominated by the alternation between the high and the low tones, obscuring the separate high and low tone patterns.

Using the same principle, that is that the relationship between two events within the same object is processed differently from that between two events located on separate objects, Müller et al. (2010) developed a paradigm for the visual modality. Participants were required to detect occasional changes of a fixation cross at the center of the screen. Irrelevant patterns composed of eight colored circles were presented around the fixation cross. From the eight circles, four complex objects were formed by connecting pairs of neighboring circles with highly visible lines. On most presentations, all circles had the same color. However, infrequently two neighboring circles were filled with a color that was different from that of the other circles (deviants). There were two kinds of deviants: the two deviant-colored circles were either connected (one-object deviant) or not (two-object deviant). The responses elicited by one-object and two-object deviants differed from each other, showing an object-related effect on vMMN.

Except for the rare case of duplex perception (Fowler and Rosenblum, 1990; Rand, 1974), each element within the sensory input belongs to exactly one object. This principle provides the basis of Rubin's (1915) well-known face–vase illusion, in which the assignment of the border between the black and the white parts of the figure determines which of them will be perceived in the foreground and thus what object is perceived in the figure. In a series of experiments, Ritter and his colleagues (De Sanctis et al., 2008; Ritter et al., 2000, 2006) demonstrated that deviant sounds elicited the MMN only with respect to the regularities present within the stream to which they belonged. Winkler et al. (2006) tested an auditory analog of Rubin's face–vase stimulus configuration and found that sounds, which could serve as the border between two concurrent sound streams, were exclusively assigned to one or the other stream.

In summary, object representations inferred from MMN and vMMN studies are separated from each other and each element of the auditory and visual input is (usually) assigned to only one of them.

2.4. Connecting temporally discrete events

This feature needs no detailed discussion. Almost all studies of MMN and vMMN delivered sequences of discrete stimuli (for reviews, see Czigler, 2007; Kimura, *in press in this issue*; Näätänen and Winkler, 1999; Näätänen et al., 2011). According to the current interpretation of the MMN (vMMN) response, the memory representations involved in the elicitation of this component encode the regularities extracted from the preceding stimulus sequence (for discussions of the alternative interpretations of MMN, see Garrido et al., 2009c; May and Tiitinen, 2010; Näätänen et al., 2005; Winkler, 2007). Thus the representations inferred from these studies connect temporally discrete stimulus events. In the auditory modality, formation of such representations has been observed with a temporal separation up to ca. 10 s between consecutive sounds (Sams et al., 1993). Other studies found that MMN was eliminated after somewhat shorter intervals (for a review, see Näätänen and Winkler, 1999). In vision, Kimura (*in press, in this issue*) reported that a small vMMN was elicited with a 5 s pre-deviant interval. Maekawa et al. (2009) tested the elicitation of vMMN while varying the duration of the interval separating the establishment of a visual regularity (by nine presentations of the same display) and the test sequence in which a deviant was delivered in the second position. No vMMN was found with 12 s separation, whereas a significant vMMN was elicited after

6 s. In a recent study (Sulykos et al, *in preparation*⁶), no vMMN was obtained after a temporal separation of 6.9 s.

Just as for the auditory MMN, it is likely that the maximal duration of the interval after which a vMMN can be elicited varies somewhat with the stimulus parameters. However, it appears that this duration is comparable between the two modalities. The temporal limitation on episodic memory representations is compatible with sensory memory findings in experimental psychology (for a review, see Coltheart, 1984).

2.5. Extracting object representations from variable sensory input

Normally, the sensory information arriving from an object is variable. Therefore, perceptual stability requires that object representations should refer to distributions along the various features rather than to concrete feature levels. In the first part of this section, we review evidence showing that the object representations inferred from (v)MMN studies can handle some variation of the stimulus features. At the same time, the requirement of distinguishing objects limits the amount of variation that should be accepted for any given object. This results in the formation of categories. In the second part of this section, we examine category effects observed in studies of deviance detection.

2.5.1. Low-level variability – higher-level regularity

In Section 2.2., we already mentioned the studies which showed that the auditory MMN is elicited by sounds deviating in a feature that is common to most of the sounds in the sequence even when other features vary. For example, Gomes et al. (1995) varied both the frequency and the intensity of the tones in a sequence while keeping duration constant. Occasional duration deviants elicited the MMN response. Furthermore, Winkler et al. (1990) showed that when the intensity of most tones varied within a given range, rare tones whose intensity fell outside this range elicited the MMN response. Thus the range of observed variance is encoded within the auditory object representations involved in the MMN-generating process.

Higher-level features are also extracted from a variable input. Several studies showed that varying the absolute frequency levels of tone pairs while keeping either only the direction or both the direction and the relative size of the within-pair pitch step constant allows the formation of a representation against which tone pairs with a different pitch step elicit the MMN (Paavilainen et al., 1999, 2003; Saarinen et al., 1992). Similar results were obtained in newborn infants (Carral et al., 2005; Stefanics et al., 2009) as well as for tonal patterns and short melodies (Fujioka et al., 2004; Tervaniemi et al., 2001, 2006; Trainor et al., 2002). The former suggest that extracting regularities from a variable input is part of the basic configuration of the auditory system. Further studies showed that more complex sequential regularities, such as generalized versions of pitch alternation (Horváth et al., 2001) or rules connecting two different features within a single sound (Paavilainen et al., 2001) are also extracted from sequences of sounds varying in one or more features.

In vision, recent studies addressed the registration of some ecologically valid high-level regularities, facial expressions of emotions. Some of these studies presented pictures of a single model (Susac et al., 2004; 2010; Zhao and Li, 2006) or schematic drawings (Chang et al., 2010). The general finding was that infrequently presented emotional expressions amongst frequent ones elicited vMMN-like responses even when the faces themselves were task-irrelevant. All studies found deviance-related parieto-occipital negativities at about 300 ms from stimulus onset. Chang et al. (2010) and Zhao

⁶ Personal communication by I. Czigler.

and Li (2006) also observed posterior negativities in the 100–200 ms range. However, these paradigms did not separate the possible effects of low-level visual features from those of the complex structure of facial features while expressing some emotion. Furthermore, none of these studies provided a strict control of attention, which causes another problem, because emotional faces are highly attention-capturing.

In order to reduce the effects specific to the visual features of a single person, Astikainen and Hietanen (2009) presented pictures from four different models. The standard stimuli were neutral faces; deviants showed either the expression of fear or happiness. Participants performed an auditory task while the pictures were presented at the fixation point. Deviance-related posterior negativities were maximal in an earlier (150–180 ms) and a later (280–320 ms) post-stimulus latency range. Deviants also elicited an anterior positivity between 130 and 170 ms from stimulus onset. In this study, the variance in faces reduced the possible confound of low-level visual features. However, due to presenting the pictures at the fixation point, the possible attention confound remained.

Stefanics et al.'s (in preparation)⁷ participants performed a visual task with stimuli presented at the center of the visual field. Four photographs of faces (two males and two females) were simultaneously presented at four peripheral locations of the screen. Ten different models (5 males and 5 females) appeared on the pictures, making the identity of the models highly variable. Within a single display, all four models expressed either happiness or fear. In separate sequences, the probabilities of happy and fearful faces were exchanged. ERPs were compared between the same emotions when they appeared frequently (standard) or infrequently (deviant) in the sequence. Both deviant emotions elicited vMMN in the 150–220 and the 250–360 ms range. Thus, even with peripheral stimulation (participants performed a task with stimuli presented in the center of the field) and with facial identity varying substantially, rare changes in facially expressed emotions were detected by the brain.

The ERP results of all of the above-reviewed studies of deviance in facial expression of emotions are quite consistent.⁸ Importantly, the emotion-related visual mismatch responses obtained by Astikainen and Hietanen (2009) and Stefanics et al (in preparation) cannot be attributed to change in some particular low-level visual feature. Thus these results showed that structural sets of visual features are extracted from a varying set of inputs and used in processing incoming stimuli. Whether this type of processing is specific to human faces remains a question for future research.

Thus both in the auditory and in the visual modality, there is evidence that higher-level structural regularities are picked up from a sequence of varying stimulus presentations. The resulting memory traces represent both what is common within the sequence and, probably, the observed range of the variation.

2.5.2. Object representations and category effects

Many studies tested the ERPs elicited by oddballs qualitatively deviating from the standard. The so-called novelty⁹ ERP effects typically consist of the elicitation of a large N1 response (for a review, see Näätänen and Picton, 1987) which may also include an MMN and the P3a (for reviews, see Escera et al., 2000; Friedman et al., 2001; Polich, 2007). Although the traditional interpretation regards P3a as an ERP index of attention switching, some recent studies suggest that it may be a higher-level evaluation of the object itself that takes into account its significance within the entire context (Horváth

et al., 2008; Rinne et al., 2006). Thus the novelty effect may be based on a category effect. This suggestion gains support from a similar study in newborn infants (Kushnerenko et al., 2007), which showed that the neonatal ERP response to deviance generally includes an early negativity in the 50–150 ms post-stimulus interval (when deviance involves relatively large spectral change) and a following positivity peaking in the 250–350 ms interval. In addition to these ERP responses, oddballs qualitatively differing from the standard stimulus also elicited a late negativity in the 450–550 ms interval. The authors interpreted this response as showing an innate tendency to form simple perceptual categories.

Another possible categorical effect was observed by van Zuijen et al. (2004, 2005), who found that rare task-irrelevant tonal patterns deviating from the frequent ones by containing an additional tone elicited MMN in musicians. Employing an ingenious design, Ruusuvirta and colleagues (2009) showed that newborn infants are also sensitive to numeric regularities. In this study, babies were presented with a series of tonal patterns made up of four pure tones. Pure tones were either 1000 or 1500 Hz in frequency. When most patterns in the sequence contained two tones of each pitch (2:2; each possible permutation delivered with equal probability), rare homogeneous patterns (4:0) elicited a discriminative response. Thus it appears that neonates extracted a categorical type of regularity from these sequences.

Many auditory deviance-detection studies of categorical effects tested contrasts between phonemes (for reviews, see Kraus and Cheour, 2000; Näätänen, 2001). MMN effects corresponding to the category-boundary (Liberman et al., 1957) as well as to the perceptual magnet effect (Kuhl, 1991) have been observed. With many studies using cross-linguistic controls, across-category contrasts yielded higher-amplitude and/or earlier MMN responses than within-category ones (Dehaene-Lambertz, 1997; Näätänen et al., 1997; Phillips et al., 2000; Winkler et al., 1999; Ylinen et al., 2005). Further, near-prototype contrasts elicited lower-amplitude MMNs than within-category contrasts between less typical pairs having equal acoustic separation (Aaltonen et al., 1997). A cross-linguistic category-boundary study of MMN in infants showed that phoneme categories emerge between the 6th and 12th month of life (Cheour et al., 1998).

Results of some studies suggest that the lexicality and meaning of word stimuli and possibly even the grammatical correctness of phrases may modulate the MMN (for a review, see Pulvermüller and Shtyrov, 2006). However, other evidence spells caution about a possible confound of familiarity (Huotilainen et al., 2001; Jacobsen et al., 2004, 2005).

Category boundary effects were also tested for colors by comparing the vMMNs elicited by contrasts between two colors with the same vs. different verbal labels but equal physical separation. Fonteneau and Davidoff (2007) found that a small within-category deviation elicited later ERP effects than an across-category one having the same amount of physical difference. However, the responses were quite atypical for vMMN. Because Czigler et al. (2002) found vMMN only for larger color deviations, it is unclear whether or not the deviance-related responses found by Fonteneau and Davidoff (2007) should be regarded as reflecting memory-based deviance detection. However, Clifford et al. (2010) found clear vMMN for across-category but not for within-category deviants having the same physical separation. Thus it is likely that the object representations involved in visual deviance detection also encode categorical information.

These studies could not shed light on whether verbal labels influence the perceptual category boundary effect, or there are genuine categorical boundaries within the visual system, which are reflected by the verbal labels and underlie subjective experience. Whether language could affect perception is highly debated. The debate was initiated by the Whorfian hypothesis of linguistic relativity (e.g. Hunt and Agnoli, 1991). To test this possibility for color categories, Thierry et al. (2009) compared vMMNs elicited by color contrasts in native

⁷ Personal communication by I. Czigler.

⁸ For a report of a further study of emotion-related vMMN, see Kimura (in press), in this issue.

⁹ In most studies, "novel" stimuli refer to pictures of real objects or environmental sounds delivered amongst simple geometric drawings or pure tones. Thus these stimuli may not be actually novel to the participant. Rather, they fall into a different category of objects and are often more complex than the frequent stimuli.

speakers of Greek and English. Unlike in English, in Greek, the blue region of colors is divided into separate color categories, one for dark blue (*ble*) and the other for light blue (*ghalazio*). There are no category differences between the two languages for the green region. Participants in [Thierry et al.'s \(2009\)](#) study responded to occasional squares within sequences predominantly consisting of circles. In separate blue and green sequences, one shade (dark or light) was frequent and the other infrequent. In native speakers of Greek, the vMMN elicited in the blue sequences was larger than that in native English speakers. No language-related difference was observed for the green sequences. This result clearly indicates that acquired categories can influence the representations underlying visual deviance detection.

In summary, the above reviewed studies showed that auditory and visual deviance detection shows category effects, suggesting that the memory representations involved in these processes encode categorical information, thus helping the separation of perceptual objects.

2.6. Predictive memory representations

In principle, all of the phenomena observed in the study of deviance detection can be explained by assuming that predictions are created for upcoming sounds by the previously formed object representations and that the stimuli arriving from the environment are compared against these predictions. Indeed, this has been suggested by [Winkler and colleagues \(Winkler, 2007; Winkler et al., 1996, 2009\)](#) for explaining the auditory MMN. [Winkler \(2007\)](#) compared the prediction-based interpretation of MMN with the then prevailing explanation based on the hypothetical strength of the memory traces involved in the MMN-generating process ([Winkler et al., 1990](#)). He showed that, whereas the prediction-based account provides a unified explanation of all MMN results, the trace-strength account requires additional assumptions to explain some of them. Furthermore, some of these additional assumptions may contradict each other (for details, see [Winkler, 2007](#)). Here we shortly mention two types of stimulus paradigms, the deviance detection results for which are especially suggestive for the prediction-based interpretation of MMN. For further discussion of the prediction-based interpretation of MMN and vMMN, see [Bendixen et al. \(in press\)](#) and [Kimura et al. \(in press\)](#), both in this issue, respectively and [Kimura et al., 2011](#).

When sounds are omitted from a sequence delivered at a fast rate, an MMN-like response is elicited which is time-locked to the expected onset of the omitted sound ([Horváth et al., 2007; Océák et al., 2006; Shinozaki et al., 2003; Winkler et al., 2005b; Yabe et al., 1997, 1998, 2001](#)). Similarly, in the visual modality, [Czigler et al. \(2006\)](#) found a vMMN-like response to omitted stimuli. Because the response is elicited in the absence of an external stimulus, the phenomenon lends itself to a temporal expectation based interpretation. However, in both modalities, elicitation of the omission response was found to be limited to presentation rates of ca. 5 Hz and above; this limitation does not seem to apply to structured sound sequences, such as a musical rhythm ([Ladinig et al., 2009](#)). This is somewhat problematic to the predictive interpretation, especially because it is obvious that humans can produce temporal expectations for longer than ca. 200-ms long periods; e.g., synchronizing their movements with slower beats. Other explanations of this phenomenon (see, [Océák et al., 2006; Shinozaki et al., 2003; Yabe et al., 1997](#); for a possible neural mechanism, see [May and Tiitinen, 2010](#)) refer to the temporal window of integration (e.g., [Zwislöck, 1969](#); for a review, see [Cowan, 1984](#)), which can be combined with the predictive account.

MMN was also obtained by violating sequential rules linking successive sounds. For example, [Paavilainen et al. \(2007\)](#) found MMN to violations of the following rule: ‘long tones are followed by high

tones and short tones are followed by low tones’. Delivering short and long tones randomly, the pitch of each tone in the sequence was specified only by the preceding tone (for a replication and improved control, see [Bendixen et al., 2008](#)). Previously, [Horváth et al. \(2001\)](#) showed the existence of a representation for a similar ‘local’ rule of pitch alternation (‘a low tone is followed by a high tone and vice versa’) in the human auditory system. In the visual modality, [Stefanics et al. \(2011\)](#) tested the effects of violating conditional probabilities. Task-irrelevant colored dot patterns were delivered to participants at the visual periphery. By shortening every second inter-stimulus interval, stimuli were perceived in terms of pairs of dot patterns. Dots within a pair had identical color in most cases and the probabilities of the different colors were equiprobable within the sequence, overall. Infrequent dot-pattern pairs, the second stimulus of which had a different color than the first one, elicited the vMMN response (for similar results in the auditory modality, see [Paavilainen et al., 1999, 2003; Saarinen et al., 1992](#)). Finding MMN and vMMN to violations of these rules hint at predictive processing, because the rules allowed predicting some feature of a stimulus from the immediately preceding one.

Thus the results of deviance detection studies are compatible with the notion of predictive memory representations. Furthermore, [Winkler \(2007\)](#) (see also [Winkler et al., 2009](#)) argued that predictive auditory object representations play an important role in making sense of complex auditory environments (termed ‘auditory scene analysis’, see [Bregman, 1990](#)) and that the primary function of the process reflected by the MMN ERP response is to update those object representations whose predictions failed (see, also [Winkler and Czigler, 1998](#)).

However, deviance-detection phenomena can also be explained without assuming predictive processing. By storing each stimulus separately, the system could search in parallel for possible ways to fit each incoming stimulus to the previous ones. The failure of this process would then lead to the elicitation of the MMN/vMMN response. Some arguments can be put forward against this retrospective account of deviance detection. For example, two MMN responses are elicited when the incoming sound mismatches two different rules (e.g., [Winkler and Czigler, 1998; Winkler et al., 1996](#)). If representations of rules are not stored, then this result indicates a system that elicits an error signal for each of the failed attempts in trying to fit the incoming sound to the previous ones. However, in the MMN/vMMN literature, a large variety of rules have been already discovered whose violations resulted in MMN elicitation. If all of these possible ways of connecting stimuli are tested in parallel (as assumed by the retrospective account), then we should see MMN being elicited for each sound, because some of the attempts are bound to fail. Furthermore, the retrospective account would have a problem explaining the reactivation phenomena found in auditory deviance detection, whereby a previously established rule, which has become apparently inactive (i.e., no MMN is elicited by violating it) can be reactivated by a single sound matching the rule ([Cowan et al., 1993; Korzyukov et al., 2003](#); for a review, see [Winkler and Cowan, 2005](#)). The longest interval after which such reactivation has been reported was 30 s ([Winkler et al., 2002](#)), which by far exceeds the temporal capacity of memory for individual sounds (ca. 10–15 s, see [Cowan, 1984](#)). Thus it is unlikely that the reactivating sound (termed ‘reminder’) could be matched to the previous ones. On the other hand, detecting a new regularity requires three presentations of the same sound outside the focus of attention ([Cowan et al., 1993; Haenschel et al., 2005; Schröger, 1997; Kimura et al., 2006; Maekawa et al., 2009](#)) and two even if the sounds are attended ([Bendixen et al., 2007](#)). Thus it is unclear how the retrospective account could explain the observed reactivation phenomena.

Even so, we acknowledge that the results of studies of deviance detection alone do not prove that predictive processing plays an important role in perception. On the other hand, corroborating evidence

obtained with other paradigms makes the case for the notion of predictive auditory and visual memory representations much stronger (for reviews, see Bar, 2007; Cheung and Bar, in press in this issue; Bendixen et al, in press in this issue; Friston, 2005; Gregory, 1980; Hohwy, 2007; Kimura, in press in this issue; Summerfield and Egner, 2009).

3. Similarities and differences between auditory and visual object representations

In the previous section, we showed that both auditory and visual memory representations, as inferred from studies of deviance detection, possess the characteristics expected of perceptual object representations. The evidence described above painted a picture of similar representations across the two modalities. Here we briefly review the possible differences between the representations in the two modalities.

The vast majority of vMMN studies adapted designs developed for auditory research. This strategy introduces a bias toward similarity between the inferred characteristics of object representations in the two modalities. The MMN method primarily investigates temporal regularities. The temporal character of auditory modality *versus* the spatial character of visual information processing is almost a truism, a belief against which we argued in this review. It is reasonable to suggest that linking together sensory information separated in time is an essential aspect of auditory object formation. Although processing temporal order is also important in vision, its representation has not yet been linked beyond doubt with the memory involved in vMMN generation. Furthermore, results obtained using other experimental methods suggest that visual temporal processes, such as those involved in the planning and regulation of scan paths are under attentional control, and inter-saccadic integration has been associated with working memory (O'Regan and Noë, 2001).

One possibility to compare auditory and visual memory representations as inferred from MMN/vMMN results is to assess how they fit with some well-known Gestalt principles (Köhler, 1947). 1. *Similarity*. As we know, the proximal stimulation is variable even when the distal object remains constant (Brunswick, 1956). We reviewed evidence (Section 2.5.1.) showing that the auditory memory representations involved in MMN generation can absorb some variability of standard stimuli. As far as we know, no similar visual studies have been conducted yet. However, the sensitivity of vMMN to experiential categories (colors) indicates the possibility that the memory involved in the vMMN response possesses similar properties. 2. *Proximity*. Stimuli closely spaced in time, i.e., stimuli within the temporal window of integration appear to form highly cohesive units. It appears that the duration of the temporal window of integration is similar in the two modalities: ~170 ms in vision (Czigler et al., 2006) and ~170–250 ms in the auditory modality (Czigler and Winkler, 1996; Horváth et al., 2007; Oceák et al., 2006; Shinozaki et al., 2003; Yabe et al., 1997, 1998). 3. *Good continuation* may contribute to object formation. MMN/vMMN responses elicited by violations of sequential rules can be considered as reflecting sensitivity to good continuation within object formation. In the auditory modality, MMN has been recorded to violations of fairly complex sequential rules (e.g., Bendixen et al., 2008; Horváth et al., 2001; Paavilainen et al., 2007; Tervaniemi et al., 1994). So far vMMN has been only obtained for violations of rather simple sequential rules (Kimura, in press in this issue). Good continuation in vision is either a property of an object perceived at a single glance (such as a sinusoidal line), or a dynamic property (such as a sinusoidal movement trajectory). The latter can be investigated by using temporarily occluded moving objects. Although motion-related vMMN has been obtained in several studies (e.g., Kremláček et al., 2006; Pazo-Alvarez et al., 2004), the effects of occlusion have not yet been tested. One may argue that in the oddball paradigm, the blank inter-stimulus

field is a model of occlusion, but we are not convinced about the validity of this paradigm as a model of occlusion.¹⁰

In summary, due to the currently still predominant auditory bias in the study of MMN, it is too early to assess the degree of similarity between the memory representations involved in auditory and visual deviance detection. We call for studies better adapted to the specifics of visual information processing, such as, for example, testing deviations related to moving objects and theoretically more relevant visual stimulus configurations.

4. Are perceptual object representations compatible with the generative models postulated by predictive coding theories?

In terms of predictive coding theories, the organism's knowledge about the world is encoded in generative models. In hierarchical predictive coding models, the system comprises nested levels with error signals propagating upwards and predictions propagating downwards. This recurrent or reciprocal message passing among levels of the hierarchical model enables the model to be optimized or adjusted; thereby selecting the best explanation for the current sensory input. No level has special relevance or carries more or less meaningful information than the others. In contrast, traditional theories of perception postulate processes and representations with specific functions and information contents, such as deviance detection and perceptual object representations, respectively. The question addressed here is whether these two general views of perception can be reconciled with each other in the case of the memory representations inferred from deviance-detection research.

Based on the reviewed evidence, we argued that the memory representations involved in auditory and visual deviance detection can be regarded as representations of perceptual objects. As such, they are the anchors grounding all further symbols manipulated in cognitive operations (Harnad, 1990). These representations are implicit (Czigler, 2007, 2010; Näätänen et al., 2011), although their information can be brought into consciousness: Many studies found close correspondence between the memory representations inferred from electrophysiological measures of deviance detection and conscious perception of the same stimuli (for reviews, see Näätänen and Winkler, 1999; Näätänen et al., 2011; Schröger, 1997, 2007). However, some studies suggest that, perhaps, not all of the information encoded in these representation can be consciously experienced (Paavilainen et al., 2001, 2007; van Zuijlen et al., 2006). Thus, perhaps they represent a stage en route toward conscious experience as suggested by Näätänen and Winkler (1999). Modality-specific perceptual object representations may also be of special relevance in the sense that they could form the basis of multi-modal perceptual object representations. However, the question remains: Can these perceptual object representations be regarded as generative models working at an intermediate level in the hierarchy postulated by predictive coding theories?

We already mentioned in Section 1 that deviance detection is a basic function in the system postulated by predictive coding theories. It produces the error signal that initiates adaptation at higher levels in the system. We also argued (Section 2.6) that deviance detection is based on comparing the incoming stimulus with the prediction from the representations of detected regularities (see also, Bendixen et al, in press in this issue; Kimura, in press in this issue; Schröger, 2007; Winkler, 2007, 2010; Winkler et al., 1996, 2009). Thus if MMN/vMMN is regarded as an error signal at some intermediate level of the predictive coding hierarchy, then the representations inferred from the MMN/vMMN studies may be related to the corresponding generative models. Indeed, Garrido et al.'s (2008, 2009a,

¹⁰ Within the sequential stimulation used in most ERP paradigms, it is difficult to separate the Gestalt law *closure* from *good continuation*.

2009b, 2009c) work showed that the scalp-recorded MMN response can be modeled in terms of a free-energy based predictive coding system. Thus it is plausible to suggest that the perceptual object representations involved in deviance detection are closely related to generative models at a specific level of the hierarchy within the predictive coding description of perception.

Such a convergence of the two frameworks of explaining perceptual processes can provide benefits for both. For example, the predictive coding view offers an immediate reconciliation between the two major hypotheses regarding the function of the MMN (vMMN) response. Traditionally, MMN was assumed to reflect a process initiating a call for further processing of the deviant stimulus as it may carry potentially important new information (Escera et al., 2000; Näätänen, 1986, 1990; Schröger, 1997). The more recent interpretation of MMN suggests that it reflects a process updating the representation of the regularities whose predictions were mismatched by the incoming sound (Winkler, 2007, 2010). However, in terms of predictive coding theories, these two functions are one and the same; in the sense that prediction errors necessarily entail an updating of representations based on novel information that the current models have failed to explain. Another important issue relates to the neural mechanisms underlying the MMN/vMMN response. Neurons throughout the ascending auditory pathway show stimulus-specific adaptation (SSA; Anderson et al., 2009; Malmierca et al., 2009; Ulanovsky et al., 2003). Based on currently available evidence, although this mechanism may feed into the deviance detection reflected by MMN, it cannot explain a large part of the ERP findings (Winkler et al., 2009). Other models attempt to explain specific MMN phenomena (Javitt et al., 1996; May and Tiitinen, 2010; Näätänen, 1984), but they miss a common framework. Predictive coding theories may fill this hole while also providing advanced modeling tools for testing hypotheses about the brain mechanisms of deviance detection.

Conversely, as was already mentioned in Section 1, current predictive coding theories are somewhat underspecified when it comes to describing the actual generative models. Deviance detection studies gathered much information in this respect. Furthermore, the hypothesis that the memory representations involved in deviance detection correspond to perceptual object representations provides a link to the vast psychological literature of perception and memory. These information can constrain predicting coding theories and further specify the models assumed by it. Finally, predictive coding theories often assume that all relevant models preexist. However, this is likely not the case: models must be built on the fly. Again, some relevant information has already been obtained by deviance-detection studies (e.g., Bendixen et al., 2007; Haenschel et al., 2005; Cowan et al., 1993; Sussman and Winkler, 2001; Winkler et al., 1996) and related models (Mill et al., 2011) and the MMN and vMMN response may provide a way to test the formation of the generative models of predictive coding theories.

5. Summary

We reviewed evidence suggesting that the memory representations involved in auditory and visual deviance detection meet the criteria set for perceptual object representations. We discussed the similarities and differences between these representations in the two modalities. Finally, we hypothesized that the memory representations involved in deviance detection are closely related to the generative models assumed by predictive coding theories.

Acknowledgments

This work was supported by the European Community's Seventh Framework Programme FP7 (Challenge 2 – Cognitive Systems, Interaction, Robotics) under grant agreement 231168-SCANDLE (to I.W.) and the Hungarian National Research Fund (OTKA) under grant agreement 71600 (to I.C.).

References

- Aaltonen, O., Eerola, O., Hellstrom, A., Uusipaikka, E., Lang, A.H., 1997. Perceptual magnet effect in the light of behavioral and psychophysiological data. *Journal of the Acoustical Society of America* 101, 1090–1105.
- Ahissar, M., Hochstein, S., 2004. The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences* 8, 457–464.
- Alain, C., Winkler, I., in press. Auditory scene analysis in the human brain: Evidence from neuroelectric recording. In Poeppl, D., Overath, T., Popper, A.N., Fay, R.R. (Eds.), *Human Auditory Cortex*, Springer Handbook of Auditory Research, Springer-Verlag, Berlin-Heidelberg-New York.
- Alink, A., Schwiedrzik, C.M., Kohler, A., Singer, W., Muckli, L., 2010. Stimulus predictability reduces responses in primary visual cortex. *Journal of Neuroscience* 30, 2960–2966.
- Anderson, L.A., Christianson, G.B., Linden, J.F., 2009. Stimulus-specific adaptation occurs in the auditory thalamus. *Journal of Neuroscience* 29, 7359–7363.
- Aoyama, A., Endo, H., Honda, S., Takeda, T., 2005. Neuromagnetic analysis of effect of audition-based prediction on visual information processing. *International Cong. Ser.* 1278, 219–222.
- Astikainen, P., Hietanen, J.K., 2009. Event-related potentials to task-irrelevant changes in facial expressions. *Behavioral and Brain Functions* 5, 30.
- Baldeweg, T., 2006. Repetition effects to sounds: evidence for predictive coding in the auditory system. *Trends in Cognitive Sciences* 10, 93–94.
- Baldeweg, T., 2007. ERP repetition effects and mismatch negativity generation: a predictive coding perspective. *Journal of Psychophysiology* 21, 204–213.
- Bar, M., 2004. Visual objects in context. *Nature Reviews Neuroscience* 7, 617–629.
- Bar, M., 2007. The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences* 11, 280–289.
- Bendixen, A., Roeber, U., Schröger, E., 2007. Regularity extraction and application in dynamic auditory stimulus sequences. *Journal of Cognitive Neuroscience* 19, 1664–1677.
- Bendixen, A., Prinz, W., Horváth, J., Trujillo-Barreto, N.J., Schröger, E., 2008. Rapid extraction of auditory feature contingencies. *NeuroImage* 41, 1111–1119.
- Bendixen, A., SanMiguel, I., Schröger, E., in press. Early electrophysiological indicators for predictive processing in audition. *International Journal of Psychophysiology*.
- Berti, S., 2011. The attentional blink demonstrates automatic deviance processing in vision. *NeuroReport* 14, 664–667.
- Boemio, A., Fromm, S., Braun, A., Poeppel, D., 2005. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience* 8, 389–395.
- Bregman, A.S., 1978. Auditory streaming: competition among alternative organizations. *Perception & Psychophysics* 23, 391–398.
- Bregman, A.S., 1990. *Auditory Scene Analysis. The Perceptual Organization of Sound*. MIT Press, Cambridge, MA.
- Bregman, A.S., Campbell, J., 1971. Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology* 89, 244–249.
- Brunswick, E., 1956. *Perception and the Representative Design of Psychological Experiments*. University of California Press, Berkeley.
- Bubic, A., von Cramon, D.Y., Schubotz, R.I., 2010. Prediction, cognition and the brain. *Front. Human Neurosci.* 4, 25.
- Carral, V., Huotilainen, M., Ruusuvirta, T., Fellman, V., Näätänen, R., Escera, C., 2005. A kind of auditory 'primitive intelligence' already present at birth. *European Journal of Neuroscience* 21, 3201–3204.
- Chang, X., Xu, J., Shi, N., Zang, B., Zhao, L., 2010. Dysfunction of processing task-irrelevant emotional faces in major depressive disorder patients revealed by expression-related visual MMN. *Neuroscience Letters* 472, 33–37.
- Cheour, M., Čeponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., Näätänen, R., 1998. Development of language specific phoneme representations in the infant brain. *Nature Neuroscience* 1, 351–353.
- Cheung, O.S., Bar, M., in press. Visual prediction and perceptual expertise. *International Journal of Psychophysiology*.
- Clifford, A., Holmes, A., Davies, I.R.L., Franklin, A., 2010. Color categories affect pre-attentive color perception. *Biological Psychology* 85, 275–282.
- Coltheart, M., 1984. Sensory memory – a tutorial review. In: Bouman, H., Bouwhuis, D.G. (Eds.), *Attention and Performance X: Control of Language Processes*. Erlbaum, Hillsdale, N.J., pp. 259–285.
- Cowan, N., 1984. On short and long auditory stores. *Psychological Bulletin* 96, 341–370.
- Cowan, N., Winkler, I., Teder, W., Näätänen, R., 1993. Short- and long-term prerequisites of the mismatch negativity in the auditory event related potential (ERP). *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19, 909–921.
- Creutzig, F., Sprekeler, H., 2008. Predictive coding and the slowness principle: an information-theoretic approach. *Neural Computation* 20, 1026–1041.
- Creutzig, F., Globerson, A., Tishby, N., 2009. Past-future information bottleneck in dynamical systems. *Physical Review E* 79.
- Czigler, I., 2007. Visual mismatch negativity – violation of nonattended environmental regularities. *Journal of Psychophysiology* 21, 224–230.
- Czigler, I., 2010. Representation or regularities in visual stimulation: event-related potentials reveal the automatic acquisition. In: Czigler, I., Winkler, I. (Eds.), *Unconscious Memory Representations in Perception: Processes and Mechanisms in the Brain*. John Benjamins, Amsterdam and Philadelphia, pp. 107–132.
- Czigler, I., Winkler, I., 1996. Preattentive auditory change detection relies on unitary sensory memory representation. *NeuroReport* 7, 2413–2417.
- Czigler, I., Balázs, L., Winkler, I., 2002. Memory-based detection of task-related visual changes. *Psychophysiology* 39, 1–5.
- Czigler, I., Winkler, I., Pató, L., Várnagy, A., Wiesz, J., Balázs, L., 2006. Visual temporal window of interaction as revealed by the mismatch negativity event-related potential to stimulus omission. *Brain Research* 1104, 129–140.

- Dayan, P., Hinton, G.E., Neal, R.M., Zemel, R.S., 1995. The Helmholtz machine. *Neural Computation* 7, 889–904.
- De Sanctis, P., Ritter, W., Molholm, S., Kelly, S.P., Foxe, J.J., 2008. Auditory scene analysis: the interaction of stimulation rate and frequency separation on pre-attentive grouping. *European Journal of Neuroscience* 27, 1271–1276.
- Deacon, D., Noursak, J.M., Pilotti, M., Ritter, W., Yang, C.-M., 1998. Automatic change detection: does the auditory system use representations of individual stimulus features or gestalts? *Psychophysiology* 35, 413–419.
- Dehaene-Lambertz, G., 1997. Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport* 8, 919–924.
- den Ouden, H.E.M., Daunizeau, J., Roiser, J., Friston, K.J., Stephan, K.E., 2010. Striatal prediction error modulates cortical coupling. *Journal of Neuroscience* 30, 3210–3219.
- Dubnov, S., 2008. Unified view of prediction and repetition structure in audio signals with application to interest point detection. *IEEE Transactions on Audio, Speech, and Language Processing* 16, 327–337.
- Duncan, J., 1984. Selective attention and the organization of visual information. *Journal of Experimental Psychology. General* 113, 501–517.
- Duncan, J., Humphreys, G.W., 1989. Visual search and stimulus similarity. *Psychological Review* 96, 433–458.
- Ernst, M.O., Banks, M.S., 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433.
- Escera, C., Alho, K., Schröger, E., Winkler, I., 2000. Involuntary attention and distractibility as evaluated with event related brain potentials. *Audiology and Neuro-Otology* 5, 151–166.
- Fonteneau, E., Davidoff, J., 2007. Neural correlated of color categories. *NeuroReport* 18, 1223–1227.
- Fowler, C.A., Rosenblum, L.D., 1990. Duplex perception: a comparison of monosyllables and slamming doors. *Journal of Experimental Psychology. Human Perception and Performance* 16, 742–754.
- Friedman, D., Cycowicz, Y.M., Gaeta, H., 2001. The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Biobehavioral Reviews* 25, 355–373.
- Friston, K., 2005. A theory of cortical responses. *Philosophical Transactions of the Royal Society London, Series B, Biological Sciences* 360, 815–836.
- Friston, K., 2010. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience* 11, 127–138.
- Friston, K., Kiebel, S., 2009. Cortical circuits for perceptual inference. *Neur. Net.* 22, 1093–1104.
- Fujioka, T., Trainor, L.J., Ross, B., Kakigi, R., Pantev, C., 2004. Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience* 16, 1010–1021.
- Garrido, M.I., Friston, K.J., Kiebel, S.J., Stephan, K.E., Baldeweg, T., Kilner, J.M., 2008. The functional anatomy of the MMN: a DCM study of the roving paradigm. *NeuroImage* 42, 936–944.
- Garrido, M.I., Kilner, J.M., Kiebel, S.J., Friston, K.J., 2009a. Dynamic causal modeling of the response to frequency deviants. *Journal of Neurophysiology* 101, 2620–2631.
- Garrido, M.I., Kilner, J.M., Kiebel, S.J., Stephan, K.E., Baldeweg, T., Friston, K.J., 2009b. Repetition suppression and plasticity in the human brain. *NeuroImage* 48, 269–279.
- Garrido, M.I., Kilner, J.M., Stephan, K.E., Friston, K.J., 2009c. The mismatch negativity: a review of underlying mechanisms. *Clinical Neurophysiology* 120, 453–463.
- Gibson, J.J., 1977. The theory of affordances. In: Shaw, R., Bransford, J. (Eds.), *Perceiving, Acting and Knowing*. Erlbaum, Hillsdale, NJ, pp. 67–82.
- Gomes, H., Ritter, W., Vaughan Jr., H.G., 1995. The nature of pre-attentive storage in the auditory system. *Journal of Cognitive Neuroscience* 7, 81–94.
- Gomes, H., Bernstein, R., Ritter, W., Vaughan Jr., H.G., Miller, J., 1997. Storage of feature conjunctions in transient auditory memory. *Psychophysiology* 34, 712–716.
- Gordon, I.E., 1997. *Theories of Visual Perception*, 2nd edn. John Wiley & Sons, New York. 171 pp.
- Gregory, R.L., 1980. Perceptions as hypotheses. *Philosophical Transactions of the Royal Society London, Series B, Biological Sciences* 290, 181–197.
- Griffiths, T.D., Warren, J.D., 2004. Opinion: what is an auditory object? *Nature Reviews Neuroscience* 5, 887–892.
- Grill-Spector, K., Henson, R., Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences* 10, 14–23.
- Haenschel, C., Vernon, D.J., Dwivedi, P., Gruzelier, J.H., Baldeweg, T., 2005. Event-related brain potential correlates of human auditory sensory memory-trace formation. *Journal of Neuroscience* 25, 10494–10501.
- Harnad, S., 1990. The symbol grounding problem. *Physica D* 42, 335–346.
- Helmholtz, H., 1860/1962. *Handbuch der Physiologischen Optik*. In: Southall, J.P.C. (Ed.), *English Translation*, Vol. 3, Dover, New York.
- Heslenfeld, D.J., 2003. Visual mismatch negativity. In: Polich, J. (Ed.), *Detection of Change: Event-related Potential and fMRI Findings*. Kluwer Academic Press, Boston, pp. 41–59.
- Hohwy, J., 2007. Functional integration and the mind. *Synthese* 159, 315–328.
- Hohwy, J., Roepstorff, A., Friston, K., 2008. Predictive coding explains binocular rivalry: an epistemological review. *Cognition* 108, 687–701.
- Holcombe, A.O., Cavanagh, P., 2001. Early binding of feature pairs for visual perception. *Nature Neuroscience* 4, 127–128.
- Hommel, B., Müsseler, G., Aschersleben, G., Prinz, W., 2001. The theory of event-coding (TEC): a framework for perception and action planning. *The Behavioral and Brain Sciences* 24, 849–878.
- Horváth, J., Czigler, I., Sussman, E., Winkler, I., 2001. Simultaneously active pre-attentive representations of local and global rules for sound sequences. *Cognitive Brain Research* 12, 131–144.
- Horváth, J., Czigler, I., Winkler, I., Teder-Sälejärvi, W.A., 2007. The temporal window of integration in elderly and young adults. *Neurobiology of Aging* 28, 964–975.
- Horváth, J., Winkler, I., Bendixen, A., 2008. DoN1/MMN, P3a, and RON form a strongly coupled chain reflecting the three stages of auditory distraction? *Biological Psychology* 79, 139–147.
- Hosoya, T., Baccus, S.A., Meister, M., 2005. Dynamic predictive coding by the retina. *Nature* 436, 71–77.
- Houck, M.R., Hoffman, J.E., 1986. Conjunction of color and form without attention: evidence from an orientation-contingent color after effect. *Journal of Experimental Psychology. Human Perception and Performance* 12, 186–199.
- Hunt, E., Agnoli, F., 1991. The Whorfian hypothesis: a cognitive psychology perspective. *Psychological Review* 98, 377–389.
- Huotilainen, M., Ilmoniemi, R.J., Lavikainen, J., Tiitinen, H., Alho, K., Sinkkonen, J., Knuutila, J., Näätänen, R., 1993. Interaction between representations of different features of auditory sensory memory. *NeuroReport* 4, 1279–1281.
- Huotilainen, M., Kujala, A., Alku, P., 2001. Long-term memory traces facilitate short-term memory trace formation in audition in humans. *Neuroscience Letters* 310, 133–136.
- Jacobsen, T., Horváth, J., Schröger, E., Lattner, S., Widmann, A., Winkler, I., 2004. Pre-attentive auditory processing of lexicality. *Brain and Language* 88, 54–67.
- Jacobsen, T., Schröger, E., Winkler, I., Horváth, J., 2005. Familiarity affects the processing of task-irrelevant ignored sounds. *Journal of Cognitive Neuroscience* 17, 1704–1713.
- Javitt, D.C., Steinschneider, M., Schroeder, C.E., Arezzo, J.C., 1996. Role of cortical N-methyl-D-aspartate receptors in auditory sensory memory and mismatch negativity generation: implications for schizophrenia. *Proceedings of the National Academy of Sciences of the USA* 93, 11962–11967.
- Jehee, J.F.M., Ballard, D.H., 2009. Predictive feedback can account for biphasic responses in the lateral geniculate nucleus. *PLoS Computational Biology* 5, e1000373.
- Kahneman, D., Treisman, A.M., Gibbs, B., 1992. The reviewing object files: object-specific integration of information. *Cognitive Psychology* 24, 175–219.
- Kersten, D., Mamassian, P., Yuille, A., 2004. Object perception as Bayesian inference. *Annals Review Psychology* 55, 271–304.
- Kimura, M., in press. Visual mismatch negativity and unintentional temporal-context-based prediction in vision. *International Journal of Psychophysiology*
- Kimura, M., Katayama, J., Murohashi, H., 2006. Probability-independent and -dependent ERPs reflecting visual change detection. *Psychophysiology* 43, 180–189.
- Kimura, M., Schröger, E., Czigler, I., 2011. Visual mismatch negativity and its importance in visual cognitive sciences. *NeuroReport* 14, 669–673.
- Knill, D., Pouget, A., 2004. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences* 27, 712–719.
- Köhler, W., 1947. *Gestalt Psychology*, 2nd edn. Liveright, New York.
- Korzyukov, O.A., Winkler, I., Gumenyuk, V.I., Alho, K., 2003. Processing abstract auditory features in the human auditory cortex. *NeuroImage* 20, 2245–2258.
- Kotz, S.A., Schwartze, M., 2010. Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends in Cognitive Sciences* 14, 392–399.
- Kraus, N., Cheour, M., 2000. Speech sound representation in the brain. *Audiology Neurology-Otology* 5, 140–150.
- Kremláček, J., Kuba, M., Kubová, Z., Langrová, J., 2006. Visual mismatch negativity elicited by magnocellular system activation. *Vision Research* 46, 485–490.
- Kubovy, M., Van Valkenburg, D., 2001. Auditory and visual objects. *Cognitive* 80, 97–126.
- Kuhl, P.K., 1991. Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* 50, 93–107.
- Kumar, S., Sedley, W., Nourski, K.V., Kawasaki, H., Oya, H., Patterson, R.D., Howard III, M.A., Friston, K.J., Griffiths, T.D., in press. Predictive coding and pitch processing in the auditory cortex. *J. Cogn. Neurosci.*
- Kushnerenko, E., Winkler, I., Horváth, J., Näätänen, R., Pavlov, I., Fellman, V., Huotilainen, M., 2007. Processing acoustic change and novelty in newborn infants. *European Journal of Neuroscience* 26, 265–274.
- Ladining, O., Honing, H., Häden, G., Winkler, I., 2009. Probing attentive and pre-attentive emergent meter in adult listeners without extensive music training. *Music Perception* 26, 377–386.
- Lee, T.S., Mumford, D., 2003. Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A* 20, 1434–1448.
- Liberman, A.M., Harris, K.S., Hoffman, H.S., Griffith, B.C., 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54, 358–368.
- Maekawa, T., Tobimatsu, S., Ogata, K., Onitsuka, T., Kanba, S., 2009. Preattentive visual change detection as reflected by the mismatch negativity (MMN) – evidence for a memory-based process. *Neurosciences Research* 65, 107–112.
- Malmierca, M.S., Cristaudo, S., Pérez-González, D., Covey, E., 2009. Stimulus-specific adaptation in the inferior colliculus of the anaesthetized rat. *Journal of Neuroscience* 29, 5483–5493.
- May, P.J.C., Tiitinen, H., 2010. Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology* 47, 66–122.
- Mill, R., Bohm, T., Bendixen, A., Winkler, I., Denham, S.L., 2011. CHAINS – competition and cooperation between fragmentary event predictors in a model of auditory scene analysis. *Information Sciences and Systems (CISS), 2011 45th Annual Conference on*. doi:10.1109/CISS.2011.5766095.
- Müller, D., Winkler, I., Roeber, U., Schaffer, S., Czigler, I., Schröger, E., 2010. Visual object representations can be formed outside the focus of voluntary attention: evidence from event-related brain potentials. *Journal of Cognitive Neuroscience* 22, 1179–1188.
- Mumford, D., 1992. On the computational architecture of the neocortex II. The role of cortico-cortical loops. *Biological Cybernetics* 66, 241–251.

- Murray, S.O., Kersten, D., Olshausen, B.A., Schrater, P., Woods, D.L., 2002. Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America* 99, 15164–15169.
- Näätänen, R., 1984. In search of a short duration memory trace of a stimulus in the human brain. In: Pulkkinen, L., Lyytinen, P. (Eds.), *Human Action and Personality, Essays in Honor of Martti Takala, Jyväskylä Studies in Education, Psychology and Social Research* 54. University of Jyväskylä, Jyväskylä, pp. 29–43.
- Näätänen, R., 1986. The orienting response theory: an integration of informational and energetical aspects of brain function. In: Hockey, R.G.J., Gaillard, A.W.K., Coles, M. (Eds.), *Adaptation to Stress and Task Demands: Energetical Aspects of Human Information Processing*. Martinus Nijhoff, Dordrecht, pp. 91–111.
- Näätänen, R., 1990. The role of attention in auditory information processing as revealed by event related potentials and other brain measures of cognitive function. *The Behavioral and Brain Sciences* 13, 201–288.
- Näätänen, R., 2001. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 38, 1–21.
- Näätänen, R., Picton, T.W., 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24, 375–425.
- Näätänen, R., Winkler, I., 1999. The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin* 125, 826–859.
- Näätänen, R., Gaillard, A.W.K., Mäntysalo, S., 1978. Early selective attention effect on evoked potential reinterpreted. *Acta Psychologica* 42, 313–329.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour Luhtanen, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., Alho, K., 1997. Language specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–434.
- Näätänen, R., Pakarinen, S., Rinne, T., Takegata, R., 2004. The mismatch negativity (MMN) – towards the optimal paradigm. *Clinical Neurophysiology* 2004, 140–144.
- Näätänen, R., Jacobsen, T., Winkler, I., 2005. Memory based or afferent processes in mismatch negativity (MMN): a review of the evidence. *Psychophysiology* 42, 25–32.
- Näätänen, R., Kujala, T., Winkler, I., 2011. Auditory processing that leads to conscious perception: a unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology* 48, 4–22.
- Neisser, U., 1967. *Cognitive Psychology*. Appleton-Century-Crofts, New York.
- O'Regan, J.K., Noë, A., 2001. A sensorimotor approach to vision and visual consciousness. *The Behavioral and Brain Sciences* 24, 883–975.
- Oceák, A., Winkler, I., Sussman, E., Alho, K., 2006. Loudness summation and the mismatch negativity event related brain potential in humans. *Psychophysiology* 43, 13–20.
- Oceák, A., Winkler, I., Sussman, E., 2008. Units of sound representation and temporal integration: a mismatch negativity study. *Neuroscience Letters* 436, 85–89.
- Paavilainen, P., Jaramillo, M., Näätänen, R., Winkler, I., 1999. Neuronal populations in the human brain extracting invariant relationships from acoustic variance. *Neuroscience Letters* 265, 179–182.
- Paavilainen, P., Simola, J., Jaramillo, M., Näätänen, R., Winkler, I., 2001. Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). *Psychophysiology* 38, 359–365.
- Paavilainen, P., Degerman, A., Takegata, R., Winkler, I., 2003. Spectral and temporal stimulus characteristics in the processing of abstract auditory features. *NeuroReport* 14, 715–718.
- Paavilainen, P., Arajärvi, P., Takegata, R., 2007. Preattentive detection of nonsalient contingencies between auditory features. *NeuroReport* 18, 159–163.
- Pakarinen, S., Huotilainen, M., Näätänen, R., 2010. The mismatch negativity (MMN) with no standard stimulus. *Clinical Neurophysiology* 121, 1043–1050.
- Pazo-Alvarez, P., Amenedo, E., Cadaveira, F., 2004. Automatic detection of motion direction changes in the human brain. *European Journal of Neuroscience* 19, 1978–1986.
- Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., Roberts, T., 2000. Auditory cortex accesses phonological categories: an MEG mismatch study. *Journal of Cognitive Neuroscience* 12, 1038–1055.
- Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication* 41, 245–255.
- Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology* 118, 2128–2148.
- Prinz, W., 1990. A common coding approach to perception and action. In: Neuman, O., Prinz, W. (Eds.), *Relationships between Perception and Action: Current Approaches*. Springer, Berlin, pp. 167–201.
- Pulvermüller, F., Shtyrov, Y., 2006. Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology* 79, 49–71.
- Rand, T.C., 1974. Dichotic release from masking for speech. *Journal of the Acoustical Society of America* 55, 678–680.
- Rao, R., Ballard, D., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience* 2, 79–87.
- Rinne, T., Sarkka, A., Degerman, A., Schröger, E., Alho, K., 2006. Two separate mechanisms underlie auditory change detection and involuntary control of attention. *Brain Research* 1077, 135–143.
- Ritter, W., Deacon, D., Gomes, H., Javitt, D.C., Vaughan Jr., H.G., 1995. The mismatch negativity of event-related potentials as a probe of transient memory: a review. *Ear and Hearing* 16, 52–67.
- Ritter, W., Sussman, E., Molholm, S., 2000. Evidence that the mismatch negativity system works on the basis of objects. *NeuroReport* 11, 61–63.
- Ritter, W., De Sanctis, P., Molholm, S., Javitt, D.C., Foxe, J.J., 2006. Preattentively grouped tones do not elicit MMN with respect to each other. *Psychophysiology* 43, 423–430.
- Rubin, E., 1915. *Synoplevede Figurer*. Gyldendalske, Copenhagen.
- Saariainen, J., Paavilainen, P., Schröger, E., Tervaniemi, M., Näätänen, R., 1992. Representation of abstract attributes of auditory stimuli in the human brain. *NeuroReport* 3, 1149–1151.
- Sams, M., Hari, R., Rif, J., Knuutila, J., 1993. The human auditory sensory memory trace persists about 10 s: Neuromagnetic evidence. *Journal of Cognitive Neuroscience* 5, 363–370.
- Schröger, E., 1997. On the detection of auditory deviations: a pre-attentive activation model. *Psychophysiology* 34, 245–257.
- Schröger, E., 2007. Mismatch negativity – a microphone into auditory memory. *Journal of Psychophysiology* 21, 138–146.
- Schütz-Bosbach, S., Prinz, W., 2007. Prospective coding in event representation. *Cognitive Proceedings* 8, 93–102.
- Shinozaki, N., Yabe, H., Sato, Y., Hiruma, T., Sutoh, T., Matsuoka, T., Kaneko, S., 2003. Spectrotemporal window of integration of auditory information in the human brain. *Cognitive Brain Research* 17, 563–571.
- Sinkkonen, J., 1999. Information and resource allocation. In: Baddeley, R., Hancock, P., Foldiak, P. (Eds.), *Information Theory and the Brain*. Cambridge University Press, Cambridge, pp. 241–254.
- Spence, C., Driver, J., 1999. Cross-modal attention. In: Humphreys, G.W., Treisman, A. (Eds.), *Attention, Space, and Action*. Oxford University Press, New York, pp. 130–149.
- Stefanics, G., Háden, G.P., Sziller, I., Balázs, L., Beke, A., Winkler, I., 2009. Newborn infants process pitch intervals. *Clinical Neurophysiology* 120, 304–308.
- Stefanics, G., Kimura, M., Czigler, I., 2011. Visual mismatch negativity reveals automatic detection of sequential regularity violation. *Frontiers in Human Neuroscience* 5, 46.
- Stefanics, G., Csukly, G., Komlósi, S., Czobor, P., Czigler, I., in preparation. Processing of unattended facial emotions.
- Sulykos, I., Czigler, I., 2011. One plus one is less than two: visual features elicit non-additive mismatch related brain activity. *Brain Research* 1398, 64–71.
- Sulykos, I., Kovács, K., Czigler, I., in preparation. Memory mismatch in vision: no reactivation.
- Summerfield, C., Egner, T., 2009. Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences* 13, 403–409.
- Susac, A., Ilmoniemi, R.J., Pihko, E., Supek, S., 2004. Neurodynamic studies on emotional and inverted faces in an oddball paradigm. *Brain Topography* 16, 265–268.
- Susac, A., Ilmoniemi, R.J., Pihko, E., Ranken, D., Supek, S., 2010. Early cortical responses are sensitive to changes in face stimuli. *Brain Research* 1346, 155–164.
- Sussman, E., 2007. A new view on the MMN and attention debate: auditory context effects. *Journal of Psychophysiology* 21, 164–175.
- Sussman, E., Winkler, I., 2001. Dynamic sensory updating in the auditory system. *Cognitive Brain Research* 12, 431–439.
- Sussman, E., Gomes, H., Nousak, J.M.K., Ritter, W., Vaughan Jr., H.G., 1998a. Feature conjunctions and auditory sensory memory. *Brain Research* 793, 95–102.
- Sussman, E., Ritter, W., Vaughan Jr., H.G., 1998b. Attention affects the organization of auditory input associated with the mismatch negativity system. *Brain Research* 789, 130–138.
- Takegata, R., Paavilainen, P., Näätänen, R., Winkler, I., 1999. Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity. *Neuroscience Letters* 266, 109–112.
- Takegata, R., Huotilainen, M., Rinne, T., Näätänen, R., Winkler, I., 2001. Changes in acoustic features and their conjunctions are processed by separate neuronal populations. *NeuroReport* 12, 525–529.
- Takegata, R., Brattico, E., Tervaniemi, M., Varyagina, O., Näätänen, R., Winkler, I., 2005. Pre-attentive representation of feature conjunctions for simultaneous, spatially distributed auditory objects. *Cognitive Brain Research* 25, 169–179.
- Tales, A., Newton, P., Troscianko, T., Butler, S., 1999. Mismatch negativity in the visual modality. *NeuroReport* 10, 3363–3367.
- Tervaniemi, M., Maury, S., Näätänen, R., 1994. Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *NeuroReport* 5, 844–846.
- Tervaniemi, M., Rytönen, M., Schröger, E., Ilmoniemi, R.J., Näätänen, R., 2001. Superior formation of cortical memory traces for melodic patterns in musicians. *Learning and Memory* 8, 295–300.
- Tervaniemi, M., Castaneda, A., Knoll, M., Uther, M., 2006. Sound processing in amateur musicians and nonmusicians: event-related potential and behavioral indices. *NeuroReport* 17, 1225–1228.
- Thierry, G., Athanasopoulos, P., Wiggert, A., Dering, B., Kuipers, J.-R., 2009. Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences of the USA* 106, 4567–4570.
- Thompson, W.F., Hall, M.D., Pressing, J., 2001. Illusory conjunctions of pitch and duration in unfamiliar tone sequences. *Journal of Experimental Psychology: Human Perception and Performance* 27, 128–140.
- Tishby, N., Polani, D., 2011. Information theory of decisions and actions. In: Cutsuridis, V., Hussain, A., Taylor, J.G. (Eds.), *Perception–Action Cycle: Models, Architectures, and Hardware*. Springer, New York, pp. 601–636.
- Trainor, L.J., McDonald, K.L., Alain, C., 2002. Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *Journal of Cognitive Neuroscience* 14, 1–13.
- Treisman, A., 1998. Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society London, Series B, Biological Sciences* 353, 1295–1306.

- Treisman, A.M., Gelade, G., 1980. A feature-integration theory of attention. *Cognitive Psychology* 12, 97–136.
- Treisman, A., Schmidt, N., 1992. Illusory conjunctions in the perception of objects. *Cognitive Psychology* 14, 107–141.
- Ulanovsky, N., Las, L., Nelken, I., 2003. Processing of low-probability sounds by cortical neurons. *Nature Neuroscience* 6, 391–398.
- van Zuijen, T.L., Sussman, E., Winkler, I., Näätänen, R., Tervaniemi, M., 2004. Pre-attentive grouping of sequential sounds an event related potential study comparing musicians and non musicians. *Journal of Cognitive Neuroscience* 16, 331–338.
- van Zuijen, T.L., Sussman, E., Winkler, I., Näätänen, R., Tervaniemi, M., 2005. Auditory organization of sound sequences by a temporal or numerical regularity: a mismatch negativity study comparing musicians and non-musicians. *Cognitive Brain Research* 23, 270–276.
- van Zuijen, T.L., Simoens, V.L., Paavilainen, P., Näätänen, R., Tervaniemi, M., 2006. Implicit, intuitive, and explicit knowledge of abstract regularities in a sound sequence: an event-related brain potential study. *Journal of Cognitive Neuroscience* 18, 1292–1303.
- Wang, W., Jones, H.E., Andolina, I.M., Salt, T.E., Sillito, A.M., 2006. Functional alignment of feedback effects from visual cortex to thalamus. *Nature Neuroscience* 9, 1330–1336.
- Weiss, Y., Simoncelli, E.P., Adelson, E.H., 2002. Motion illusions as optimal percepts. *Nature Neuroscience* 5, 598–604.
- Wightman, F.L., Jenison, R., 1995. Auditory spatial layout. In: Epstein, W., Rogers, S.J. (Eds.), *Perception of Space and Motion*, 2nd edn. Academic Press, San Diego, CA, pp. 365–400.
- Winkler, I., 2007. Interpreting the mismatch negativity (MMN). *Journal of Psychophysiology* 21, 147–163.
- Winkler, I., 2010. In search for auditory object representations. In: Czigler, I., Winkler, I. (Eds.), *Unconscious Memory Representations in Perception: Processes and Mechanisms in the Brain*. John Benjamins, Amsterdam and Philadelphia, pp. 71–106.
- Winkler, I., Cowan, N., 2005. From sensory to long-term memory – evidence from auditory memory reactivation studies. *Experimental Psychology* 52, 3–20.
- Winkler, I., Czigler, I., 1998. Mismatch negativity: deviance detection or the maintenance of “standard”. *NeuroReport* 9, 3809–3813.
- Winkler, I., Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., Näätänen, R., 1990. The effect of small variation of the frequent auditory stimulus on the event related brain potential to the infrequent stimulus. *Psychophysiology* 27, 228–235.
- Winkler, I., Karmos, G., Näätänen, R., 1996. Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event related potential. *Brain Research* 742, 239–252.
- Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., Aaltonen, O., Raimo, I., Alho, K., Lang, H., Iivonen, A., Näätänen, R., 1999. Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Brain Research. Cognitive Brain Research* 7, 357–369.
- Winkler, I., Korzyukov, O., Gumenyuk, V., Cowan, N., Linkenkaer-Hansen, K., Alho, K., Ilmoniemi, R.J., Näätänen, R., 2002. Temporary and longer retention of acoustic information. *Psychophysiology* 39, 530–534.
- Winkler, I., Sussman, E., Tervaniemi, M., Ritter, W., Horváth, J., Näätänen, R., 2003. Pre-attentive auditory context effects. *Cognitive, Affective, and Behavioral Neuroscience* 3, 57–77.
- Winkler, I., Czigler, I., Sussman, E., Horváth, J., Balázs, L., 2005a. Preattentive binding of auditory and visual stimulus features. *Journal of Cognitive Neuroscience* 17, 320–339.
- Winkler, I., Takegata, R., Sussman, E., 2005b. Event-related brain potentials reveal multiple stages in the perceptual organization of sound. *Cognitive Brain Research* 25, 291–299.
- Winkler, I., van Zuijen, T., Sussman, E., Horváth, J., Näätänen, R., 2006. Object representation in the human auditory system. *European Journal of Neuroscience* 24, 625–634.
- Winkler, I., Denham, S.L., Nelken, I., 2009. Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences* 13, 532–540.
- Yabe, H., Tervaniemi, M., Reinikainen, K., Näätänen, R., 1997. Temporal window of integration revealed by MMN to sound omission. *NeuroReport* 8, 1971–1974.
- Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilmoniemi, R.J., Näätänen, R., 1998. The temporal window of integration of auditory information in the human brain. *Psychophysiology* 35, 615–619.
- Yabe, H., Winkler, I., Czigler, I., Koyama, S., Kakigi, R., Sutoh, T., Hiruma, T., Kaneko, S., 2001. Organizing sound sequences in the human brain: the interplay of auditory streaming and temporal integration. *Brain Research* 897, 222–227.
- Ylinen, S., Shestakova, A., Alku, P., Huotilainen, M., 2005. The perception of phonological quantity based on durational cues by native speakers, second-language users and non-speakers of Finnish. *Language and Speech* 48, 313–338.
- Yu, A.J., 2007. Adaptive behavior: humans act as Bayesian learners. *Current Biology* 17, R977–R980.
- Yuille, A., Kersten, D., 2006. Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Science* 10, 301–308.
- Zhao, L., Li, J., 2006. Visual mismatch negativity elicited by facial expressions under non-attentional condition. *Neuroscience Letters* 401, 126–131.
- Zwislocki, J.J., 1969. Temporal summation of loudness: an analysis. *Journal of the Acoustical Society of America* 46, 431–440.